

RESEARCH ARTICLE

Open Access



# Patient survival and tumor characteristics associated with *CHEK2*:p.I157T – findings from the Breast Cancer Association Consortium

Taru A. Muranen<sup>1</sup>, Carl Blomqvist<sup>2</sup>, Thilo Dörk<sup>5</sup>, Anna Jakubowska<sup>6</sup>, Päivi Heikkilä<sup>4</sup>, Rainer Fagerholm<sup>1</sup>, Dario Greco<sup>7</sup>, Kristiina Aittomäki<sup>3</sup>, Stig E. Bojesen<sup>8,9,10</sup>, Mitul Shah<sup>11</sup>, Alison M. Dunning<sup>11</sup>, Valerie Rhenius<sup>11</sup>, Per Hall<sup>13</sup>, Kamila Czene<sup>13</sup>, Judith S. Brand<sup>13</sup>, Hatfeg Darabi<sup>13</sup>, Jenny Chang-Claude<sup>15,16</sup>, Anja Rudolph<sup>15</sup>, Børge G. Nordestgaard<sup>8,10</sup>, Fergus J. Couch<sup>17</sup>, Steven N. Hart<sup>18</sup>, Jonine Figueroa<sup>19</sup>, Montserrat García-Closas<sup>19,20</sup>, Peter A. Fasching<sup>21,22</sup>, Matthias W. Beckmann<sup>21</sup>, Jingmei Li<sup>13</sup>, Jianjun Liu<sup>23</sup>, Irene L. Andrulis<sup>24,25</sup>, Robert Winqvist<sup>26,27</sup>, Katri Pylkäs<sup>26,27</sup>, Arto Mannermaa<sup>28,29,30</sup>, Vesa Kataja<sup>28,31</sup>, Annika Lindblom<sup>13</sup>, Sara Margolin<sup>14</sup>, Jan Lubinski<sup>6</sup>, Natalia Dubrowskaja<sup>5</sup>, Manjeet K. Bolla<sup>12</sup>, Joe Dennis<sup>12</sup>, Kyriaki Michailidou<sup>12,32</sup>, Qin Wang<sup>12</sup>, Douglas F. Easton<sup>11,12</sup>, Paul D. P. Pharoah<sup>11,12</sup>, Marjanka K. Schmidt<sup>33</sup> and Heli Nevanlinna<sup>1\*</sup>

## Abstract

**Background:** P.I157T is a *CHEK2* missense mutation associated with a modest increase in breast cancer risk. Previously, another *CHEK2* mutation, the protein truncating c.1100delC has been associated with poor prognosis of breast cancer patients. Here, we have investigated patient survival and characteristics of breast tumors of germ line p.I157T carriers.

**Methods:** We included in the analyses 26,801 European female breast cancer patients from 15 studies participating in the Breast Cancer Association Consortium. We analyzed the association between p.I157T and the clinico-pathological breast cancer characteristics by comparing the p.I157T carrier tumors to non-carrier and c.1100delC carrier tumors. Similarly, we investigated the p.I157T associated risk of early death, breast cancer-associated death, distant metastasis, locoregional relapse and second breast cancer using Cox proportional hazards models. Additionally, we explored the p.I157T-associated genomic gene expression profile using data from breast tumors of 183 Finnish female breast cancer patients (ten p.I157T carriers) (GEO: GSE24450). Differential gene expression analysis was performed using a moderated *t* test. Functional enrichment was investigated using the DAVID functional annotation tool and gene set enrichment analysis (GSEA). The tumors were classified into molecular subtypes according to the St Gallen 2013 criteria and the PAM50 gene expression signature.

(Continued on next page)

\* Correspondence: Heli.Nevanlinna@hus.fi

<sup>1</sup>Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, P.O. Box 70000029, HUS, Helsinki, Finland  
Full list of author information is available at the end of the article



© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

**Results:** p.I157T was not associated with increased risk of early death, breast cancer-associated death or distant metastasis relapse, and there was a significant difference in prognosis associated with the two *CHEK2* mutations, p.I157T and c.1100delC. Furthermore, p.I157T was associated with lobular histological type and clinico-pathological markers of good prognosis, such as ER and PR expression, low TP53 expression and low grade. Gene expression analysis suggested luminal A to be the most common subtype for p.I157T carriers and CDH1 (cadherin 1) target genes to be significantly enriched among genes, whose expression differed between p.I157T and non-carrier tumors.

**Conclusions:** Our analyses suggest that there are fundamental differences in breast tumors of *CHEK2*:p.I157T and c.1100delC carriers. The poor prognosis associated with c.1100delC cannot be generalized to other *CHEK2* mutations.

**Keywords:** Breast cancer, CHEK2, CHK2, I157T, 1100delC, Survival, Pathology, Gene expression

## Background

Checkpoint kinase 2 (*CHEK2*) is a moderate penetrance breast cancer risk gene. The two most frequent *CHEK2* mutations in European populations are p.I157T and c.1100delC. Truncating *CHEK2* founder mutations (c.1100delC, IVS2 + 1G > A, del5395) confer a higher than twofold increase in the risk of breast cancer [1–3], whereas p.I157T (c.470 T > C, rs17879961), a *CHEK2* missense mutation is associated with a milder, 1.4-fold elevation in the risk [4]. The c.1100delC carrier frequency is highest in the Netherlands and in Finland (over 1 %), the other two truncating founder mutations are found mainly in Poland [3], and p.I157T is most frequent in Finland and in Poland (around 5 %) [5]. Additionally, dozens of rare *CHEK2* missense mutations have been found in breast cancer patients, but their contribution to disease risk is minor on a population level and causative role in disease development probably varies greatly [6–8].

The consequences of c.1100delC and p.I157T differ on a molecular level, but both have been shown to severely interfere with the CHEK2 protein activity. C.1100delC is a loss-of-function mutation that induces a premature termination codon in the kinase domain in exon 10 (ter381) leading to a nonsense-mediated mRNA decay, which reduces both mutated and overall *CHEK2* mRNA level [9, 10]. C.1100delC truncates CHEK2 protein's C-terminal kinase domain. The truncated protein is unstable and practically undetectable in mutation carrier cells [9]. Isoleucine 157 (p.I157T) is required for several van der Waals interactions at the interface of forkhead-associated (FHA) and kinase domains of dimerizing CHEK2 peptide chains. Its replacement to threonine (p.I157T) has been shown to interfere with these interactions and to severely impede the CHEK2 homodimerization required for its activation [11]. Furthermore, ectopic expression of human *CHEK2*:p.I157T failed a *rad53/sml* complementation assay in yeast suggesting an impaired protein function [6]. Thus, p.I157T possibly disturbs CHEK2 function by competing with the wild-type protein

in dimer formation in heterozygous cells in a dominant negative manner [4].

Since both p.I157T and c.1100delC cause increased risk of breast cancer and compromise the activity of the CHEK2 protein, the question remains whether their effects on patient prognosis would be proportional to their risk effects and how similar the breast cancer phenotypes associated with the mutations would be. C.1100delC is associated with bilateral disease and estrogen receptor (ER)-positive tumors [12–14]. However, although tumors from p.I157T carriers are also predominantly ER-positive [15], tumors from p.I157T and c.1100delC carriers are associated with phenotypically different types of breast cancer. The lobular histological type is overrepresented among p.I157T mutation carrier tumors [16], whereas the c.1100delC carrier tumors are typically ductal [13, 14].

We have previously reported *CHEK2*:c.1100delC heterozygosity to be associated with reduced overall and disease-free survival as well as with increased risk of breast cancer-specific death in a Breast Cancer Association Consortium (BCAC) data set combining mutation carriers from multiple European populations [17]. Here, we report a study investigating thoroughly the prognostic associations of *CHEK2*:p.I157T as well as pathologic characteristics and genomic gene expression profiles of breast tumors from carriers of germ line p.I157T.

## Methods

### Study subjects for survival and pathology analyses

We included in the analyses female invasive breast cancer patients of European ancestry with a first invasive primary breast cancer enrolled in 15 studies participating in the Breast Cancer Association Consortium (BCAC) (Additional file 1: Table S1). In order to be able to stratify the analyses by study, only BCAC studies providing genotype and survival data of about ten *CHEK2*:p.I157T carriers were included in the analyses (Additional file 1: Table S2). Altogether, the data set consisted of 26,801 study subjects, of which 590 carried germ line p.I157T and 271 carried c.1100delC mutations (Table 1).

**Table 1** Tumor characteristics of the BCAC study subjects

		Non-carriers	p.1157T carriers	c.1100delC carriers	p value (1157T/nc)	p value (1157T/1100delC)	p value (1100delC/nc)
ER	Negative	4595	85	26	0.00046	0.61	0.0015
		20.2 %	15.5 %	11.2 %			
	Positive	18,179	462	207			
		79.8 %	84.5 %	88.8 %			
	Missing	3166	43	38			
		12.2 %	7.3 %	14.0 %			
PR	Negative	6397	147	44	0.0034	0.90	0.0045
		32.7 %	28.5 %	23.3 %			
	Positive	13,173	368	145			
		67.3 %	71.5 %	76.7 %			
	Missing	6370	75	82			
		24.6 %	12.7 %	30.3 %			
Her2	Negative	8220	231	95	0.68	0.10	0.24
		84.7 %	83.7 %	81.9 %			
	Positive	1483	45	21			
		15.3 %	16.3 %	18.1 %			
	Missing	16,237	314	155			
		62.6 %	53.2 %	57.2 %			
EGFR	Negative	3841	122	62	0.21	0.26	0.034
		89.6 %	90.4 %	96.9 %			
	Positive	448	13	2			
		10.4 %	9.6 %	3.1 %			
	Missing	21,651	455	207			
		83.5 %	77.1 %	76.4 %			
CK5/6	Negative	4734	143	80	0.30	0.29	0.19
		87.9 %	88.3 %	92.0 %			
	Positive	652	19	7			
		12.1 %	11.7 %	8.0 %			
	Missing	20,554	428	184			
		79.2 %	72.5 %	67.9 %			
TP53	Negative	3755	144	88	0.00048	0.21	0.16
		81.6 %	90.6 %	86.3 %			
	Positive	847	15	14			
		18.4 %	9.4 %	13.7 %			
	Missing	21,338	431	169			
		82.3 %	73.1 %	62.4 %			
Tumor size (ordinal)	<20 mm	14,949	340	149	0.29	0.33	0.83
		65.6 %	65.6 %	62.6 %			
	20–50 mm	6953	162	82			
		30.5 %	31.3 %	34.5 %			
	>50 mm	876	16	7			
		3.8 %	3.1 %	2.9 %			
	Missing	3162	72	33			
		12.2 %	12.2 %	12.2 %			

**Table 1** Tumor characteristics of the BCAC study subjects (*Continued*)

Lymph node status	Negative	13,144	320	125	0.94	0.91	0.64
		62.0 %	60.4 %	57.6 %			
	Positive	8070	210	92			
		38.0 %	39.6 %	42.4 %			
	Missing	4726	60	54			
Grade (ordinal)	1	4916	132	38	0.00023	0.0030	0.38
		22.5 %	26.7 %	17.4 %			
	2	10,817	266	127			
		49.6 %	53.7 %	58.0 %			
	3	6089	97	54			
		27.9 %	19.6 %	24.7 %			
	Missing	4118	95	52			
		15.9 %	16.1 %	19.2 %			
	Ductal	14,133	273	193			
		72.7 %	60.0 %	76.9 %			
Histological type	Lobular	2966	100	36	0.0044*	0.0010 <sup>#</sup>	0.67 <sup>#</sup>
		15.3 %	22.0 %	14.3 %			
	Mixed (ductal and lobular)	455	26	4			
		2.5 %	6.0 %	1.7 %			
	Tubular	271	17	2			
		1.5 %	4.0 %	0.7 %			
	Medullary	177	4	3			
		1.0 %	0.9 %	1.7 %			
	Mucinous	213	9	3			
		1.2 %	2.1 %	1.4 %			
	Papillary	55	1	1			
		0.3 %	0.2 %	1.8 %			
	Missing†	7670	160	29			
		29.6 %	27.1 %	10.7 %			
Subtype‡	LumA (ER+, PR+, Her2-)	5415	164	72	0.00089	0.46	0.0087
		58.9 %	63.6 %	65.5 %			
	LumB (ER+, PR-, Her2- or ER+, Her2+)	1939	62	30			
		21.1 %	24.0 %	27.3 %			
	Basal (ER-, PR-, Her2-)	1306	16	4			
		14.2 %	6.2 %	3.6 %			
	Her2-positive (ER-, PR-, Her2+)	536	16	4			
		5.8 %	6.2 %	3.6 %			
	Missing	16,744	332	161			
		64.5 %	56.3 %	59.4 %			
AGE cat (ordinal)	50 or younger	6932	162	101	0.99	0.0037	0.0029
		27.0 %	27.8 %	37.3 %			
	Older than 50 and not more than 70	16,083	344	151			
		62.6 %	59.0 %	55.7 %			
	Older than 70	2674	77	19			
		10.4 %	13.2 %	7.0 %			

**Table 1** Tumor characteristics of the BCAC study subjects (Continued)

	Missing	251	7	0			
		1.0 %	1.2 %	0.0 %			
AGE	Mean		57.1	57.9	54.3	0.41	0.028
St.dev.	0.024	10.8	11.1	11.0			
Total		25,940	590	271			

BCAC Breast Cancer Association Consortium, ER estrogen receptor, PR progesterone receptor, Her2 human epidermal growth factor receptor 2, EGFR epidermal growth factor receptor, CK5/6 cytokeratin 5/6, TP53 tumor protein 53, LumA luminal A, LumB luminal B

<sup>a</sup>Categories Medullary, Mucinous and Papillary were combined for the Cochran-Mantel-Haenszel test

<sup>#</sup>Categories Mixed, Tubular, Medullary, Mucinous and Papillary were combined for the Cochran-Mantel-Haenszel test

<sup>†</sup>The "missing" category included also rare forms of breast cancer, which did not belong to the named categories: 1179 non-carriers, 25 p.I157T carriers and 9 c.1100delC carriers

<sup>\*</sup>Tumor subtypes are defined according to ER, PR and Her2 expression following the St Gallen 2013 guidelines [34]

Italics is used to indicate the proportion of study subjects in each category. E.G. 'ER-positive/all with known ER-status' or 'missing/all study subjects'

Individuals carrying both mutations were excluded from the analyses ( $n = 4$ ).

### Genotyping

*CHEK2*:p.I157T was first genotyped by independent studies using various methods including MassARRAY iPLEX Gold (Sequenom, San Diego, CA, USA), TaqMan (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) and Fluidigm (Fluidigm, San Francisco, CA, USA) as listed in Additional file 1: Table S1. Quality control was implemented as follows: each study performed duplicate measurements of at least two samples from each sample plate and genotyped 93 CEPH control DNAs (HAPMAPPT01, Coriell Institute for Medical Research, Cambden, NJ, USA). If a study reported more than two discordant genotyping results of the CEPH DNAs, all genotype data from that study was excluded. Later, p.I157T was genotyped centrally using a custom Illumina iSelect genotyping array for the Collaborative Oncological Gene-environment Study (COGS) [18]. Discordant genotyping results were clarified with Sanger sequencing. *CHEK2*:c.1100delC was genotyped by independent studies using mainly TaqMan (Additional file 1: Table S1), as described earlier [17].

### Pathology analysis

Pathology data was collected from hospital records or from scientific projects within the individual studies, as described previously [19]. Additionally, the TP53 protein expression was measured by individual studies using immunohistochemical staining as described in Additional file 1: Table S3. The pathology data availability and mutation carrier frequencies varied between independent BCAC studies and therefore all analyses were stratified by study. Pathology analyses were performed using R environment for statistical computing version 3.0.2 [20] including packages vcdExtra [21] and meta [22]. Comparisons were made between *CHEK2* mutations carriers (heterozygous or homozygous) and non-carriers, for both p.I157T and c.1100delC, as well as between carriers of p.I157T and c.1100delC (Table 1). Associations between the mutations and clinico-

pathological characteristics were tested with study-stratified Cochran-Mantel-Haenszel test (mantelhaen.test for categorical characteristics and CMHtest for ordinal characteristics). The category of missing data was not included in these comparisons. Differences in age at diagnosis were tested by meta-analysis of age distribution in independent studies using a random effects model (metacont).

### Survival analysis

Survival analyses were performed using the Cox regression [23] as implemented in Stata (Stata/SE 10.1 for Windows, StataCorp LP, College Station, TX, USA) comparing *CHEK2* mutation carriers and non-carriers, as described above. Study subjects were considered to become at risk at the time of their first invasive breast cancer diagnosis. The data did not consist entirely of incident cases. Therefore, in order to avoid bias caused by late enrollment, we implemented a method called left censoring, which has been proven to provide robust survival estimates for data, which includes also prevalent cases [24]. Survival analysis endpoints included death of any cause, breast cancer-associated death, distant metastasis relapse, locoregional relapse and second breast cancer. Patients were censored at the end of their follow-up period or at the latest 15 years after the initial breast cancer diagnosis in analyses of overall survival and second breast cancer, but at the latest 10 years in analyses of locoregional or distant relapse-free survival as well as in analyses of breast cancer-specific survival. Patients presenting with distant metastases at diagnosis were excluded from the analyses of locoregional relapse-free survival. All analyses were stratified by study.

In addition to univariate analyses, we performed multivariate analyses, which were stratified by study and age category ( $\leq 50$  years;  $> 50$  and  $\leq 70$ ;  $> 70$ ), and adjusted for tumor grade (1, 2 or 3, ordinal), tumor size (1: maximum diameter less than or equal to 20 mm; 2: more than 20 mm and less than or equal to 50 mm; 3 over 50 mm, ordinal), tumor spread in axillary lymph nodes (0 = negative, 1 = positive) and progesterone receptor (PR)

status (0 = negative, 1 = positive). ER was not included in the model, because of the non-linear relationship between tumor ER status and patient survival during the 10 years following the diagnosis; patients with ER-negative tumors have a higher risk of dying from breast cancer during the first 5 years after the diagnosis, but the difference in risk between ER-positive and ER-negative tumors levels out after that period [17, 25]. However, since several studies have reported an association between the two *CHEK2* mutations and ER-positive disease [12–14] (Table 1), we performed the survival analyses in a subgroup of patients with ER-positive tumors. Only cases with complete data on the pathological markers were included in the multivariate analyses. Univariate survival analyses were performed also in a subgroup of breast cancer patients with lobular tumors, because of the association between p.I157T and lobular breast cancer [15] (Table 1).

### Study subjects for gene expression analysis

Gene expression analyses were performed using a data set of 183 breast tumors from the Helsinki University Hospital (GEO: GSE24450). As described previously, the data set consisted of total RNA samples from 151 tumors from unselected cohorts of breast cancer patients and 32 tumors from additional familial cases hybridized on Illumina HumanHT-12 v3 Expression BeadChips (Illumina Inc., San Diego, CA, USA) [10, 26]. The p.I157T carrier status was defined from peripheral blood samples as described earlier for the BCAC study 'HEBCS' (Additional file 1: Table S1). Ten patients were germ line p.I157T carriers and 162 were non-carriers, of which six carried germ line c.1100delC. The c.1100delC carrier tumors were included in the analyses as non-I157T carriers. The p.I157T genotype information was not available for 11 study subjects. These were included in the molecular subtype analysis, but not in differential gene expression or gene set enrichment analysis. The clinico-pathologic characteristics of the 183 tumors are provided in Additional file 1: Table S4.

### Gene expression analysis

Gene expression data quality control and quantile normalization was performed in the Bioconductor [27] as described earlier [26]. Data analyses were performed in R version 3.0.2 and Bioconductor packages *genefu* [28], *limma* [29, 30] and *geneplotter* [31]. Probes not mapping to any current Entrez Gene entities (GRCh38.p2) were excluded, resulting in a filtered data set of 20,145 genes.

For determining the intrinsic molecular subtypes, expression data of the fifty PAM50 signature genes was extracted from the filtered data set, median centered and standardized per gene by dividing with the standard deviation of the gene's expression values. Intrinsic subtypes were defined by Pearson correlation between tumors and the luminal A,

luminal B, human epidermal growth factor receptor 2 (Her2)-enriched, basal-like and normal-like centroids as implemented in the *genefu* package [28, 32]. Hierarchical clustering was performed using the Ward's method [33]. As a comparison to the subtype classification by gene expression, we used the surrogate clinico-pathologic markers to define the subtypes following the St Gallen 2013 criteria (luminal A: ER+, PR+, Her2-, Ki-67-; luminal B (three marker combinations): ER+, PR-, Her2- or ER+, Her2-, Ki-67+ or ER+, Her2+; basal: ER-, PR-, Her2-; Her2 overexpressing: ER-, PR-, Her2+) [34].

For analysis of differential gene expression the data was filtered by including only genes with highest variation in expression levels over the entire data set (st. dev.  $\geq 0.75$ , 1852 genes). The samples from p.I157T carriers were compared to samples from non-carriers with a moderated *t* test adjusting for ER, tumor protein 53 (TP53) and Ki-67 protein expression (positive/negative), tumor grade (1, 2, 3, ordinal) as well as histological type (lobular/other). The adjusting covariates were selected from features tabulated in Additional file 1: Table S4 as the most significant factors ( $p < 0.001$ ) explaining variation in the expression of the 1852 genes as summarized by the first four principal components. Additionally, lobular histologic type was included to avoid bias caused by the association between the p.I157T and lobular type. Data on at least one of the adjusting variables was missing for 12 tumor samples and thus the differential gene expression and gene set enrichment analyses were performed with a set of 160 (ten p.I157T and 150 non-carrier) tumor samples and 1852 genes. Genes with *p* values below 0.01 were considered to be associated with p.I157T. These were included in a functional enrichment analysis performed using the DAVID functional annotation tool [35]. Functional annotations with Benjamini-Hochberg [36] corrected *p* values below 0.01 were considered to be significantly enriched.

Gene set enrichment analysis (GSEA) was performed using a java application available at <http://software.broadinstitute.org/gsea> following the instructions of the user guide [37]. For the GSEA analysis, the 1852 genes were ranked according to a score calculated as the product of  $\log_2(\text{fold change})$  and  $\log_{10}(p \text{ value})$  from comparisons of p.I157T carrier and non-carrier tumors as described above. All gene sets available at the Molecular Signatures Database (MSigDB) v5.0 [38] were included in the analyses. The *p* values were corrected for false discovery rate for all other gene sets but the gene sets originating from single publications ('CGP: chemical and genetic perturbations' database), which were corrected for the family-wise error rate. Gene sets with corrected *p* value below 0.05 were considered to be significantly enriched in the p.I157T carrier tumors.



## Results and discussion

Our findings from extensive analyses of breast tumor phenotypes and patient survival underline a fundamental difference in breast cancers of the carriers of two *CHEK2* mutations, p.I157T and c.1100delC. Significant differences were found in tumor grade and histopathological type as well as in patient survival of p.I157T and c.1100delC carriers, whereas no difference was seen in tumor subtypes: ER+, PR+, Her2- disease was the most common type for carriers of both mutations.

### Association of p.I157T with clinico-pathological markers

In our analyses p.I157T was associated with low grade as well as several other markers, which have previously been associated with good prognosis (Table 1). Our analyses confirmed the previously reported associations between p.I157T and ER-positive or lobular breast cancer [15]. Also mixed (ductal and lobular) and tubular histological types were more frequent in p.I157T carrier tumors. Both of ER-positive and lobular tumor types are associated with good short-term prognosis, but increased risk of disease progression in the long run [25, 39]. Furthermore, p.I157T was associated with PR-positive and TP53-negative breast cancer. PR expression is a marker for good prognosis for ER-positive breast cancer and it has been suggested as a surrogate marker separating luminal A and luminal B subtypes according to immunohistochemical data [34, 40, 41]. TP53 immunohistochemical staining is considered to be indicative of somatic *TP53* mutations. Strong TP53 staining suggests the presence of stabilizing mutations (primarily missense), whereas absence of staining indicates typically a protein-truncating mutation (nonsense or frameshift), and weak staining a wild-type functional *TP53*. Both strong and completely negative TP53 staining have been associated with poor prognosis in comparison to weak staining [42–44]. The sensitivity of the assays used in this study did not enable differentiation between normal, low or absent TP53 expression. Therefore, we used binary classification of TP53 immunohistochemical data, the positive category corresponding to high expression (strong staining) and negative category to low expression (Additional file 1: Table S3). Noteworthy, the loss-of-function mutations associated with absent TP53 staining are relatively rare in breast cancer: these are seen in less than 5 % of all tumors [45, 46]. Therefore, it is likely that most of the tumors in the category 'negative' (Table 1) represented tumors with wild-type *TP53*. However, compromised *CHEK2* function as a result of the p.I157T mutation could be another way for TP53 silencing as *CHEK2* is among the key upstream activators of TP53 [5].

Like p.I157T, also c.1100delC was associated with ER-positive and PR-positive disease in our data set (Table 1). Furthermore, TP53-positive tumors were slightly less

often observed in c.1100delC carriers than in non-carriers, even though the difference was not statistically significant. Significant differences in clinico-pathological features associated with the two *CHEK2* mutations were seen in grade and histological type, as the c.1100delC carrier tumors resembled more non-carrier tumors (Table 1).

### Breast cancer subtypes

We investigated the I157T-associated molecular breast cancer subtypes by applying the St Gallen 2013 criteria for immunohistochemical markers [34] on the BCAC data set as well as St Gallen 2013 and the PAM50 classifier [32] on the gene expression data set of 183 breast tumors. The subtype classification of the BCAC study subjects relied on the available immunohistochemical markers, ER, PR and Her2. We found both p.I157T and c.1100delC carrier tumors to be predominantly ER+, ER+, Her2-, suggestive of good prognosis ER+ tumors or the luminal A subtype (Table 1) [34]. Also the frequency of ER+ subtypes linked to poor prognosis (ER+, PR-, Her2-; ER+, Her2+), referred to as luminal B [34], were more common for *CHEK2* mutation carriers than for non-carriers. This confirmed previous reports with regard to p.I157T [15], but was not consistent with previous reports on c.1100delC-associated tumor subtypes [47, 48]. However, the difference between our findings and these reports may have arisen from different overall cohort compositions or from differing classification methods, as the guidelines for subtype classification have changed over the years.

Subtype classifications of the 183 tumors according to gene expression data and immunohistochemical markers were partly contradictory (Table 2). Similar inconsistencies between gene expression-based classification and the surrogate immunohistochemical markers have been reported previously for other data sets [40, 49]. Overall, the division between basal and luminal appeared rather consistent: only 17 (9 %) of the 183 tumors were classified differentially across the luminal-basal axis. PAM50 [32] classified three of the p.I157T carrier tumors as luminal A, two as luminal B and two as basal. Three lobular tumors were classified as normal-like (Table 2). This kind of a misclassification has been reported to be typical for lobular tumors due to their infiltrating growth pattern, which causes the tumor sample to consist of unusually high proportion of non-cancerous stromal cells [39]. St Gallen 2013 criteria classified these normal-like tumors as luminal (ER+). Furthermore, in unsupervised hierarchical clustering of the 183 tumor samples based on expression of the PAM50 signature genes (Fig. 1), two of the normal-like p.I157T tumors (HEL\_045 and HEL\_174) clustered within the luminal A branch suggesting that luminal A could be their true molecular subtype. In summary, luminal A appeared to be the most common

**Table 2** Phenotypic classification of breast tumors from ten carriers of p.I157T

	Intrinsic subtype PAM50	IHC subtype St Gallen 2013 criteria	Histological type
HEL_045	Normal	LumA	Lobular
HEL_055	LumB	LumB	Ductal
HEL_086	LumA	LumA	Lobular
HEL_126	LumA	Basal	Ductal
HEL_128	LumA	LumB	Ductal
HEL_131	Basal	Her2pos	In situ
HEL_144	Normal	LumB	Lobular
HEL_150	Basal	LumB	Ductal
HEL_163	LumB	LumB	Ductal
HEL_174	Normal	LumA	Lobular

IHC immunohistochemistry, LumB luminal B, LumA luminal A, Her2 human epidermal growth factor receptor 2

subtype for p.I157T carrier tumors in the gene expression data concordantly with the findings in the BCAC data.

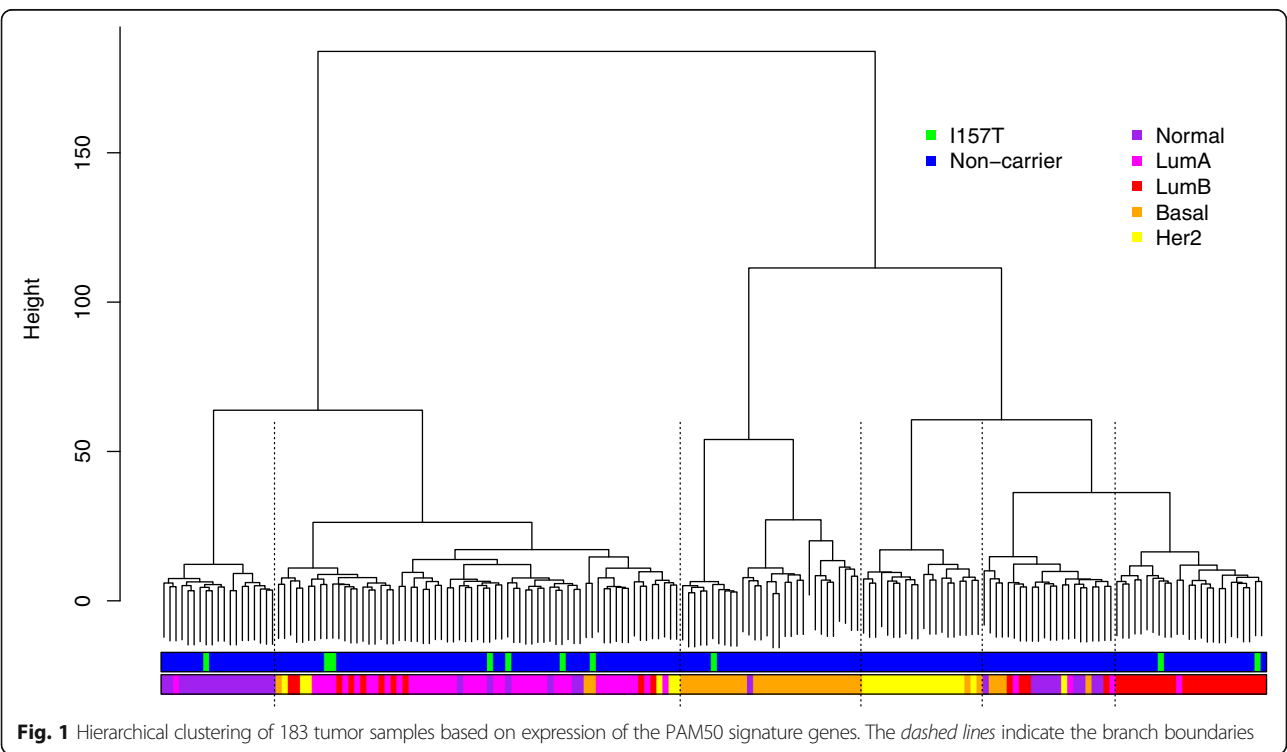
#### Patient survival

P.I157T carriers had better prognosis than the c.1100delC carriers with regard to overall or breast cancer-specific survival (Table 3a and b). This difference was possibly due to the poor survival associated with c.1100delC as reported previously by several studies [13, 14, 17, 50]. No statistically

significant difference in overall or breast cancer-specific survival was seen between p.I157T carriers and non-carriers. Hazard ratios in the analyses of subgroups of patients with ER-positive or lobular tumors were comparable to those of the main analyses (Table 3b and c).

Noteworthy, the c.1100delC carriers included here were only a subset of the study subjects included in the previous report by Weischer and colleagues on survival of c.1100delC carriers of the BCAC studies [17]. This was because the individual BCAC studies, which did not provide sufficient number of p.I157T carriers, were excluded from these analyses. Thus, the lack of statistical significance in some comparisons of survival difference between c.1100delC carriers and non-carriers (Table 3a and b) probably only reflected limited power due to low number of c.1100delC carriers, since the hazard ratios were always consistent with the previous report.

The different prognoses associated with p.I157T and c.1100delC possibly reflect their difference in molecular level severity of functional consequences. Therefore, it would be tempting to assume that the prognosis of all carriers of the truncating mutations would be similar to the prognosis of c.1100delC carriers. However, a recent Polish study combining three truncating *CHEK2* founder mutations found no difference between mutation carrier and non-carrier survival [51]. Some part of the conflicting findings could be explained by different patient selection: in the Polish study all patients had been diagnosed before





**Table 3** Risk of death or disease recurrence associated with *CHEK2*:p.I157T

(a) All breast cancer patients	Univariate analysis			Adjusted analysis		
	I157T/nc	I157T/1100delC	1100delC/nc	I157T/nc	I157T/1100delC	1100delC/nc
Early death	0.85 [0.68 - 1.07] <i>0.16</i>	0.74 [0.50 - 1.09] <i>0.12</i>	1.28 [1.00 - 1.64] <i>0.054</i>	0.80 [0.60 - 1.07] <i>0.13</i>	0.51 [0.29 - 0.90] <i>0.0190</i>	1.32 [0.94 - 1.86] <i>0.11</i>
Breast cancer-specific death	0.85 [0.60 - 1.20] <i>0.36</i>	0.64 [0.37 - 1.12] <i>0.12</i>	1.44 [1.04 - 2.00] <i>0.030</i>	0.93 [0.62 - 1.40] <i>0.73</i>	0.46 [0.21 - 1.03] <i>0.058</i>	1.25 [0.78 - 2.00] <i>0.36</i>
Distant metastasis relapse	1.04 [0.79 - 1.37] <i>0.79</i>	0.66 [0.38 - 1.14] <i>0.13</i>	1.38 [0.90 - 2.11] <i>0.14</i>	1.05 [0.75 - 1.47] <i>0.76</i>	0.62 [0.31 - 1.23] <i>0.17</i>	1.37 [0.83 - 2.26] <i>0.22</i>
Locoregional relapse	1.43 [0.92 - 2.23] <i>0.11</i>	0.81 [0.58 - 1.13] <i>0.21</i>	2.07 [1.16 - 3.69] <i>0.014</i>	1.62 [0.99 - 2.66] <i>0.056</i>	0.91 [0.33 - 2.52] <i>0.85</i>	1.26 [0.59 - 2.70] <i>0.55</i>
Second breast cancer	1.54 [0.85 - 2.78] <i>0.15</i>	0.69 [0.47 - 1.03] <i>0.070</i>	2.88 [1.68 - 4.98] <i>0.00015</i>	2.03 [1.05 - 3.92] <i>0.035</i>	0.69 [0.42 - 1.13] <i>0.14</i>	3.62 [1.82 - 7.21] <i>0.00026</i>
(b) Patients with ER+ breast cancer						
Early death	Univariate analysis			Adjusted analysis		
	I157T/nc	I157T/1100delC	1100delC/nc	I157T/nc	I157T/1100delC	1100delC/nc
Early death	0.81 [0.61 - 1.07] <i>0.14</i>	0.62 [0.39 - 0.99] <i>0.044</i>	1.32 [0.98 - 1.78] <i>0.067</i>	0.77 [0.55 - 1.07] <i>0.12</i>	0.46 [0.25 - 0.85] <i>0.013</i>	1.52 [1.06 - 2.17] <i>0.022</i>
Breast cancer-specific death	0.80 [0.51 - 1.23] <i>0.30</i>	0.47 [0.23 - 0.96] <i>0.038</i>	1.46 [0.96 - 2.22] <i>0.074</i>	0.80 [0.49 - 1.32] <i>0.39</i>	0.33 [0.13 - 0.84] <i>0.019</i>	1.50 [0.92 - 2.45] <i>0.10</i>
Distant metastasis relapse	1.00 [0.71 - 1.40] <i>0.98</i>	0.55 [0.29 - 1.02] <i>0.057</i>	1.58 [0.99 - 2.54] <i>0.056</i>	1.03 [0.70 - 1.51] <i>0.88</i>	0.56 [0.26 - 1.19] <i>0.13</i>	1.61 [0.94 - 2.77] <i>0.083</i>
Locoregional relapse	1.46 [0.86 - 2.47] <i>0.16</i>	0.77 [0.52 - 1.14] <i>0.19</i>	2.33 [1.19 - 4.57] <i>0.014</i>	1.58 [0.90 - 2.79] <i>0.11</i>	0.93 [0.29 - 2.98] <i>0.90</i>	1.08 [0.44 - 2.66] <i>0.87</i>
Second breast cancer	1.33 [0.64 - 2.75] <i>0.44</i>	0.58 [0.37 - 0.92] <i>0.019</i>	4.09 [2.31 - 7.26] <i>1.4E-06</i>	1.81 [0.82 - 3.96] <i>0.14</i>	0.61 [0.36 - 1.04] <i>0.067</i>	4.39 [2.17 - 8.87] <i>3.8E-05</i>
(c) Patients with lobular breast cancer						
Early death	Univariate analysis					
	I157T/nc					
Early death	0.67 [0.39 - 1.15] <i>0.14</i>					
Breast cancer-specific death	0.91 [0.46 - 1.80] <i>0.79</i>					
Distant metastasis relapse	0.87 [0.48 - 1.57] <i>0.64</i>					
Locoregional relapse	2.45 [0.95 - 6.34] <i>0.065</i>					
Second breast cancer	1.92 [0.57 - 6.49] <i>0.29</i>					

Hazard ratios with 95 % confidence intervals (in parenthesis) and *p* values (italics) are reported from comparisons of p.I157T carriers and non-carriers (nc) as well as comparisons of p.I157T carriers and c.1100delC carriers. All analyses were stratified by study. Multivariate analyses were stratified by study and age category, and adjusted for tumor grade, size, progesterone receptor and nodal status. Analyses were performed also in subgroups of (b) patients with estrogen receptor-positive tumors and (c) patients with lobular tumors  
ER estrogen receptor

50 years of age, whereas here and in Weischer et al. [17] also postmenopausal patients were included in the analyses. Another potential explanation could be mutation-specific survival effects. As the Polish study combined in the analyses three different truncating mutations, the c.1100delC specific effects could have been masked, since

it is the least common of the three truncating *CHEK2* mutations in Polish population [52]. Similarly as here, the Polish study reported no significant difference in survival of the p.I157T carriers and non-carriers [51].

The hazard ratios for locoregional relapse and second breast cancer (91 % contralateral, 9 % ipsilateral)

associated with p.I157T and c.1100delC were close to the mutations' relative risk estimates of primary breast cancer (Table 3) [5, 17, 53]. The marginally significant increased risk of locoregional relapse associated with p.I157T in the adjusted analyses (hazard ratio 1.62 [0.99 - 2.66], *p* value 0.056) warrants further studies, but could merely reflect the baseline risk associated with p.I157T: some of the local recurrences could represent new cancers arising during the 10-year follow-up. The risk of locoregional relapse for c.1100delC carriers was elevated in the univariate analysis but leveled out in the adjusted analysis.

#### **p.I157T associated differentially expressed genes**

In order to investigate the molecular biology of p.I157T carrier tumors and to identify potential tumor-driving events and pathways, we performed an analysis of differential gene expression and subsequent functional enrichment analysis comparing ten p.I157T to 150 non-carrier tumors. We found 21 genes to be differentially expressed between p.I157T and non-carrier tumors. All of these had higher expression in the p.I157T carrier tumors (Table 4). When the 160 tumor samples were clustered according to expression of these 21 genes, the p.I157T

tumors did not form a distinct cluster (Fig. 2), suggesting that high expression of these genes is not exclusive of the p.I157T mutation carrier tumors, but typical for a subgroup of breast tumors including the mutation carrier tumors. Tumors with different intrinsic subtypes appeared to be dispersed across all branches, similarly as the c.1100delC carriers suggesting that in this data set the c.1100delC carrier tumors would not be similar to the p.I157T tumors.

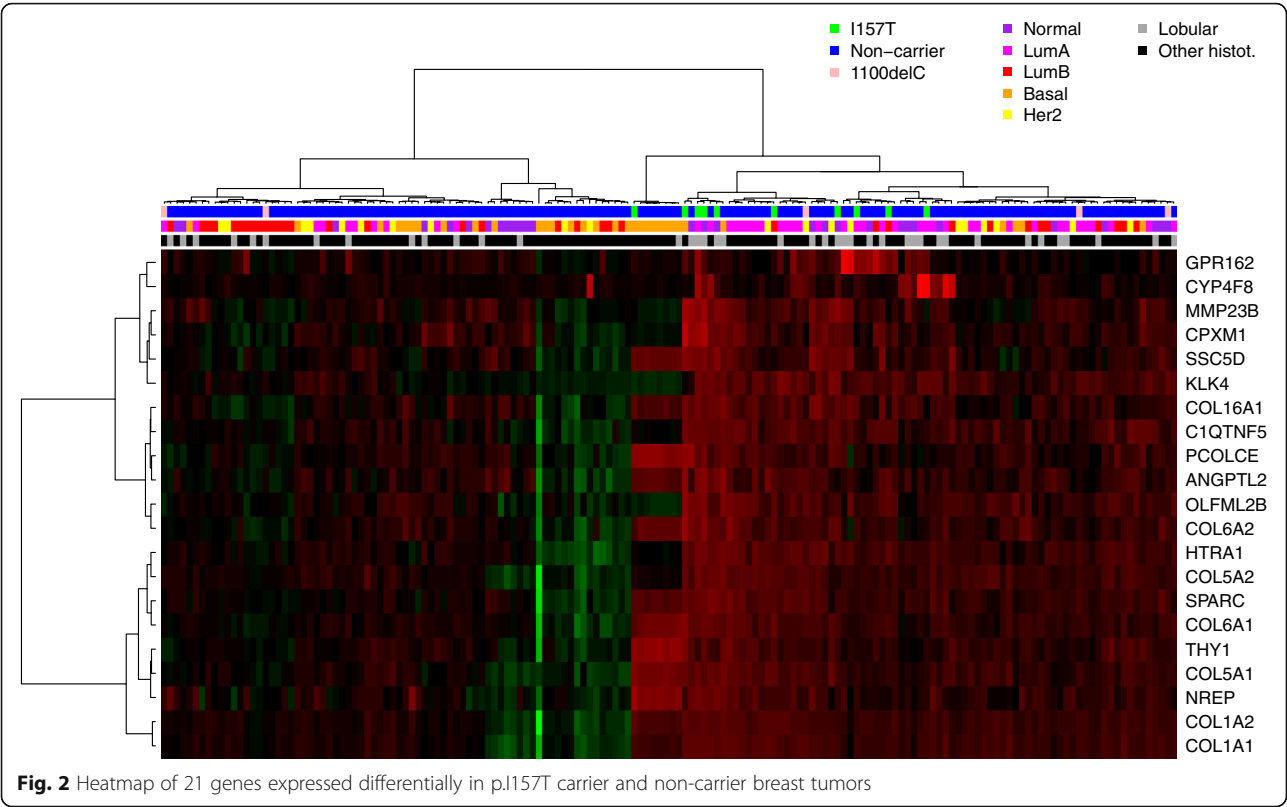
#### **Enrichment of features associated with lobular breast cancer**

The list of 21 differentially expressed genes contained seven collagen genes (Table 4), which were a major driver in the functional enrichment analysis. The enriched annotations from DAVID [54] analysis included characteristics of the collagen family and their related functions such as 'focal adhesion,' 'extracellular matrix (ECM) organization' and 'ECM-receptor interaction' (Additional file 1: Table S5). Similar results were obtained from the GSEA analysis (Additional file 1: Table S6, Additional file 2: Figure S1). Since collagens are usually expressed by stromal fibroblasts, the findings may suggest that infiltrating growth pattern, typical for lobular tumors [39], could be more

**Table 4** Differentially expressed genes in breast tumors of p.I157T carriers when compared to non-carrier tumors

Gene ID	Symbol	Description	logFC	<i>p</i> value
11283	CYP4F8	Cytochrome P450, family 4, subfamily F, polypeptide 8	1.10	0.00026
1289	COL5A1	Collagen, type V, alpha 1	1.24	0.00049
1292	COL6A2	Collagen, type VI, alpha 2	0.90	0.00056
56265	CPXM1	Carboxypeptidase X (M14 family), member 1	1.04	0.00063
1277	COL1A1	Collagen, type I, alpha 1	1.25	0.0011
23452	ANGPTL2	Angiopoietin-like 2	0.85	0.0020
5118	PCOLCE	Procollagen C-endopeptidase enhancer	0.90	0.0021
284297	SSC5D	Scavenger receptor cysteine rich family, 5 domains	0.84	0.0025
25903	OLFML2B	Olfactomedin-like 2B	0.88	0.0026
27239	GPR162	G protein-coupled receptor 162	0.73	0.0026
5654	HTRA1	HtrA serine peptidase 1	0.96	0.0027
9315	NREP	Neuronal regeneration-related protein	0.78	0.0028
1278	COL1A2	Collagen, type I, alpha 2	1.02	0.0046
8510	MMP23B	Matrix metalloproteinase 23B	0.73	0.0055
114902	C1QTNF5	C1q and tumor necrosis factor-related protein 5	0.77	0.0057
6678	SPARC	Secreted protein, acidic, cysteine-rich (osteonectin)	0.88	0.0073
9622	KLK4	Kallikrein-related peptidase 4	1.03	0.0082
1291	COL6A1	Collagen, type VI, alpha 1	0.92	0.0084
1307	COL16A1	Collagen, type XVI, alpha 1	0.73	0.0090
1290	COL5A2	Collagen, type V, alpha 2	0.93	0.0095
7070	THY1	Thy-1 cell surface antigen	0.88	0.0098

logFC logarithm of fold change

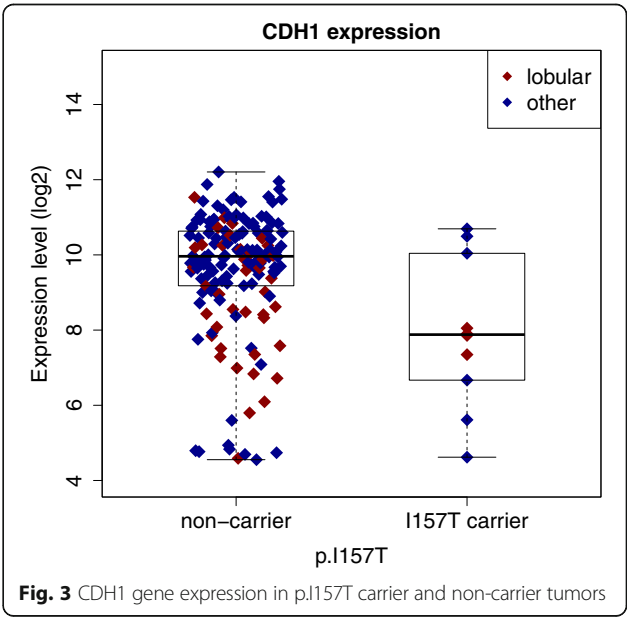


common also for non-lobular p.I157T carrier tumors than for the non-carrier tumors. Further support for this hypothesis came from the GSEA, which showed cadherin 1 (CDH1) target genes to be significantly enriched among genes, whose expression was lower in p.I157T than in non-carrier tumors (Additional file 1: Table S6). CDH1 silencing is generally considered as a defining characteristic of lobular tumors and it is often caused by somatic mutations targeting the *CDH1* gene itself [39]. However, since the differential gene expression analysis, which was also the basis for the ranked gene list used as an input to GSEA, was adjusted for the lobular tumor type, the impact of the diagnosed lobular cancers on these findings should have been minimal. *CDH1* gene expression was lower in p.I157T carriers tumors in the adjusted analysis (log<sub>2</sub> fold change -1.12, *p* value 0.03, Fig. 3), but it did not exceed the preset threshold for significance. Previously, we have reported *CDH1* mRNA expression to be higher in c.1100delC carrier than in non-carriers tumors [10]. Therefore, *CDH1* expression appears to be yet another factor, which is not shared by breast tumors from carriers of the two CHEK2 mutations, p.157 T and c.1100delC, and possibly reflects somatic changes, which have taken place during the clonal evolution of the p.I157T carrier tumors [39]. Taken together, these results suggest that besides the fact that the lobular tumors are more common among p.I157T carriers and non-carriers, the association between

p.I157T and lobular features could be even stronger than what is suggested by the diagnosed histological types.

#### Enrichment of cancer associated gene signatures

In the GSEA analysis, several independent MSigDB [38] gene signatures related to epithelial-to-mesenchymal



transition (EMT) [55–57], stromal stem cells [58] or invasive behavior [59, 60] were enriched at the top of the gene list with higher expression in p.I157T carrier tumors than in non-carrier tumors (Additional file 1: Table S6). These observations may reflect higher stromal content of the p.I157T carrier tumor samples, as the samples were not prepared at a single cell level. However, to prevent such confounding effects the tumor sample sections were selected by an experienced breast cancer pathologist. Furthermore, the above mentioned MSigDB signatures originated from carefully designed experiments tailored to detect the true signal from cancerous epithelial cells and to escape the effects of non-cancerous stromal cells. The enrichment of these signatures may suggest that the p.I157T carrier tumors have an intrinsically invasive nature. However, this should have been reflected into poor prognosis for the p.I157T carriers, which we did not see in the survival analyses. On the other hand, it is possible that higher state of differentiation of the tumor cells suggested by low grade accompanied with the invasive nature can be seen in the prognosis only in the long run, and within the 10-year follow-up period is only reflected in the slightly elevated risk of local recurrence. All in all, these observations deserve further studies before any definitive conclusions can be made.

In addition to CDH1, tumor suppressor retinoblastoma 1 (RB1) appeared as a potential gene expression regulator, whose activity was reduced in p.I157T carrier tumors in comparison to non-carrier tumors (Additional file 1: Table S6, Additional file 1: Figure S1). RB1 and its direct downstream target E2F-1 are both targets of the CHEK2 protein [61, 62]. Thus, the differential expression of the RB1 target genes possibly reflects compromised CHEK2 function in the p.I157T carrier tumors.

Noteworthy, the two differential gene expression studies on c.1100delC carrier tumors have reported enrichment of genes of WNT and FGF pathways [10, 47], which regulate the growth and differentiation of normal breast epithelium [63–66]. Among the p.I157T-associated differentially expressed genes we did not see enrichment of any growth factor pathway. These notions on differences in gene expression signatures are more descriptive than definitive by nature, but they further emphasize intrinsic biological differences between p.I157T and 1100deC carrier tumors.

## Conclusions

Based on our analyses, breast cancers of p.I157T and c.1100delC *CHEK2* mutation carriers differ in disease severity as seen especially in differences in tumor grade and patient survival, as well as in intrinsic biological features as seen in differences in histological type and gene expression profiles. Thus, it appears that even

though both mutations have been proven to compromise the protein function [6, 9, 11], they have different consequences on the disease phenotype, and prognostic findings based on one mutation cannot be generalized to the other. Furthermore, our results raise a hypothesis that the increased risk of locoregional relapse for p.I157T carriers could be caused by intrinsically invasive nature of the tumor cells. Future studies with longer follow-up are needed to test this hypothesis.

## Additional files

**Additional file 1: Table S1.** Description, genotyping methods and references for individual studies as reported by the studies. **Table S2.** Genotype and follow-up data availability in individual studies. **Table S3.** Sources and scoring of TP53 immunohistochemistry data used in this study. **Table S4.** Pathological characteristics of 183 breast tumors used in the gene expression analysis. **Table S5.** Functional annotations enriched in the 21 differentially expressed genes. **Table S6.** Enriched gene sets at the high and low edges of a list of 1852 genes ranked according to differences between p.I157T and non-carrier breast tumors. (DOCX 72 kb)

**Additional file 2: Figure S1.** Enrichment plots of the high ranking gene sets from the gene set enrichment analyses (GSEA). (PDF 364 kb)

## Abbreviations

BCAC: Breast Cancer Association Consortium; CDH1: Cadherin 1; CHEK2: Checkpoint kinase 2; ECM: Extracellular matrix; EMT: Epithelial-to-mesenchymal transition; ER: Estrogen receptor; FHA: Forkhead associated; GSEA: Gene set enrichment analysis; Her2: Human epidermal growth factor receptor 2; PR: Progesterone receptor; RB1: Retinoblastoma 1; TP53: Tumor protein 53

## Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. Especially we would like to acknowledge Kirsimari Aaltonen, Karl von Smitten, Sofia Khan, Tuomas Heikkinen, Irja Erkkilä; The SEARCH and EPIC teams; The Swedish Medical Research Counsel; Petra Seibold, Dieter Flesch-Jany, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels; staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen; Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner, Teresa Selander, Nayana Weerasooriya; Meeri Otsukka, Kari Mononen; Eija Myöhänen, Helena Kemiläinen; and Michael Bremer. The iCOGS study would not have been possible without the contributions of the following: Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai and Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog and Ken Offit (CIMBA), Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génomique Québec Innovation Centre, Sune F. Nielsen and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility.

## Funding

BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS); The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen; The HEBCCS was financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, The Nordic Cancer Union and the Sigrid Juselius Foundation. The work of TAM has been supported by Finnish Cultural Foundation,

Orion Research Foundation, Cancer Society of Finland and Ida Montin Foundation; SEARCH is funded by a programme grant from Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge; The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer; The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]; The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital; The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation; The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA; The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A\*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation; The Ontario Familial Breast Cancer Registry (OFBCR) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR; The OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities; The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland; Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and and Bert von Kantzows foundation; The SZBCS was supported by Grant PBZ\_KBN\_122/P05/2004; The HABCS study was supported by an intramural grant from Hannover Medical School; Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

#### Availability of data and materials

The dataset supporting the conclusions of this article is available in the Gene Expression Omnibus (GEO) repository, GEO: GSE24450, <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24450>.

#### Authors' contributions

TAM, HN, MKS, CB, TD, AJ and PDP conceived and designed the study, ran the data analyses and interpreted the results. HN, PDP, PH and DFE provided financial support. MKB, QW and DFE provided administrative support. All authors provided study material and patients, participated in collecting and assembling the data as well as in manuscript writing. All authors approved the final version of the manuscript.

#### Competing interests

Dr. Peter Fasching has had a consultant or advisory relationship with Roche and Novartis, has received honoraria from Novartis and Genomic Health, and has received research funding from Novartis and Amgen. Other authors declare no conflicts of interest.

#### Ethics approval and consent to participate

All participating studies were approved by their institutional review committees. All participants provided written informed consent or did not object to the secondary use of their tissue and data following country-specific regulations.

#### Author details

<sup>1</sup>Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, P.O. Box 70000029, HUS, Helsinki, Finland. <sup>2</sup>Department of Oncology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>3</sup>Department of Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>4</sup>Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>5</sup>Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. <sup>6</sup>Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. <sup>7</sup>Unit of Systems Toxicology, Finnish Institute of Occupational Health, Helsinki, Finland. <sup>8</sup>Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>9</sup>Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. <sup>10</sup>Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. <sup>11</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. <sup>12</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>13</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>14</sup>Department of Oncology - Pathology, Karolinska Institutet, Stockholm, Sweden. <sup>15</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>16</sup>University Cancer Center Hamburg (UCC), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>17</sup>Department of Laboratory Medicine and Pathology Mayo Clinic, Rochester, MN, USA. <sup>18</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. <sup>19</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. <sup>20</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. <sup>21</sup>Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany. <sup>22</sup>Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA. <sup>23</sup>Human Genetics Division, Genome Institute of Singapore, Singapore, Singapore. <sup>24</sup>Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada. <sup>25</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. <sup>26</sup>Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland. <sup>27</sup>Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit and Biocenter Oulu, University of Oulu, Oulu, Finland. <sup>28</sup>Cancer Center, Kuopio University Hospital, Kuopio, Finland. <sup>29</sup>Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. <sup>30</sup>Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. <sup>31</sup>Central Finland Hospital District, Jyväskylä Central Hospital, Jyväskylä, Finland. <sup>32</sup>Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. <sup>33</sup>Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

Received: 1 April 2016 Accepted: 15 September 2016

Published online: 03 October 2016

#### References

1. CHEK2 Breast Cancer Case-Control Consortium. CHEK2\*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet.* 2004;74(6):1175–82.
2. Johnson N, Fletcher O, Naceur-Lombardelli C, dos Santos SI, Ashworth A, Peto J. Interaction between CHEK2\*1100delC and other low-penetrance breast-cancer susceptibility genes: a familial study. *Lancet.* 2005;366(9496):1554–7.
3. Cybulski C, Wokolorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, et al. Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. *J Clin Oncol.* 2011;29(28):3747–52.
4. Kilpivaara O, Vahteristo P, Falck J, Syrjäkoski K, Eerola H, Easton D, et al. CHEK2 variant I157T may be associated with increased breast cancer risk. *Int J Cancer.* 2004;111(4):543–7.



5. Nevanlinna H, Bartek J. The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene*. 2006;25(43):5912–9.
6. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. *Hum Mol Genet*. 2012;21(12):2738–44.
7. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegelé C, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res*. 2011;13(1):R6.
8. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM, et al. Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. *Int J Cancer*. 2007;121(12):2661–7.
9. Anczukow O, Ware MD, Buisson M, Zetoune AB, Stoppa-Lyonnet D, Sinilnikova OM, et al. Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins? *Hum Mutat*. 2008;29(1):65–73.
10. Muranen TA, Greco D, Fagerholm R, Kilpivaara O, Kampjarvi K, Aittomäki K, et al. Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. *Breast Cancer Res*. 2011;13(5):R90.
11. Cai Z, Chehab NH, Pavletich NP. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol Cell*. 2009;35(6):818–29.
12. Kilpivaara O, Bartkova J, Eerola H, Syrjäkoski K, Vahteristo P, Lukas J, et al. Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. *Int J Cancer*. 2005;113(4):575–80.
13. de Bock GH, Schutte M, Krol-Warmerdam EM, Seynaeve C, Blom J, Brekelmans CT, et al. Tumor characteristics and prognosis of breast cancer patients carrying the germline CHEK2\*1100delC variant. *J Med Genet*. 2004;41(10):731–5.
14. Schmidt MK, Tollenaar RA, de Kemp SR, Broeks A, Cornelisse CJ, Smit VT, et al. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2\*1100delC germline mutation. *J Clin Oncol*. 2007;25(1):64–9.
15. Domagala P, Wokolorczyk D, Cybulski C, Huzarski T, Lubinski J, Domagala W. Different CHEK2 germline mutations are associated with distinct immunophenotypic molecular subtypes of breast cancer. *Breast Cancer Res Treat*. 2012;132(3):937–45.
16. Huzarski T, Cybulski C, Domagala W, Gronwald J, Byrski T, Swiec M, et al. Pathology of breast cancer in women with constitutional CHEK2 mutations. *Breast Cancer Res Treat*. 2005;90(2):187–9.
17. Weischer M, Nordestgaard BG, Pharoah P, Bolla MK, Nevanlinna H, Van't Veer LJ, et al. CHEK2\*1100delC heterozygosity in women with breast cancer associated with early death, breast cancer-specific death, and increased risk of a second breast cancer. *J Clin Oncol*. 2012;30(35):4308–16.
18. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353–61.
19. Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*. 2011;20(16):3289–303.
20. R Development Core Team. R: A language and environment for statistical computing. 2012.
21. Friendly M. vcdExtra: 'vcd' extensions and additions. 2015. R package version 0.6-8.
22. Schwarzer G. meta: General package for meta-analysis. R package version 4.1-0. 2015.
23. Cox DR. Regression models and life-tables. *J Roy Statist Soc Ser B*. 1972;34:187–220.
24. Azzato EM, Greenberg D, Shah M, Blows F, Driver KE, Caporaso NE, et al. Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? *Br J Cancer*. 2009;100(11):1806–11.
25. Colzani E, Liljegren A, Johansson AL, Adolfsson J, Hellborg H, Hall PF, et al. Prognosis of patients with breast cancer: causes of death and effects of time since diagnosis, age, and tumor characteristics. *J Clin Oncol*. 2011;29(30):4014–21.
26. Heikkinen T, Greco D, Pelttari LM, Tommiska J, Vahteristo P, Heikkilä P, et al. Variants on the promoter region of PTEN affect breast cancer progression and patient survival. *Breast Cancer Res*. 2011;13(6):R130.
27. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
28. Haibe-Kains B, Schroeder M, Bontempi G, Sotiriou C, Quackenbush J. geneFu: Relevant functions for gene expression analysis, especially in breast cancer. 2013. R package version 1.12.0.
29. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer; 2005.
30. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics*. 2007;23(20):2700–7.
31. Gentleman R, Bioconductor. geneplotter: Graphics related functions for Bioconductor; R package version 1.40.0.
32. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–7.
33. Ward Jr JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 1963;58:236–44.
34. Curigliano G, Criscitiello C, Andre F, Colleoni M, Di Leo A. Highlights from the 13th St Gallen International Breast Cancer Conference 2013. Access to innovation for patients with breast cancer: how to speed it up? *Eccancermedicalsociety*. 2013;7:299.
35. Anonymous. DAVID Bioinformatics Database. 2010. <http://david.abcc.ncifcrf.gov/home.jsp>.
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57(1):289–300.
37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
38. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–40.
39. McCart Reed AE, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics'. *Breast Cancer Res*. 2015;17:12.
40. Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *J Clin Oncol*. 2013;31(2):203–9.
41. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res*. 2007;9(1):R6.
42. Yemelyanova A, Vang R, Kshirsagar M, Lu D, Marks MA, Shih I, et al. Immunohistochemical staining patterns of p53 can serve as a surrogate marker for TP53 mutations in ovarian carcinoma: an immunohistochemical and nucleotide sequencing analysis. *Mod Pathol*. 2011;24(9):1248–53.
43. Watanabe G, Ishida T, Furuta A, Takahashi S, Watanabe M, Nakata H, et al. Combined immunohistochemistry of PLK1, p21, and p53 for predicting TP53 status: an independent prognostic factor of breast cancer. *Am J Surg Pathol*. 2015;39(8):1026–34.
44. Valgardsdottir R, Tryggvadottir L, Steinarsdottir M, Olafsdottir K, Jonasdottir S, Jonasson JG, et al. Genomic instability and poor prognosis associated with abnormal TP53 in breast carcinomas. *Molecular and immunohistochemical analysis*. *APMIS*. 1997;105(2):121–30.
45. Olivier M, Langerod A, Carrieri P, Bergh J, Klaar S, Eyfjord J, et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clin Cancer Res*. 2006;12(4):1157–67.
46. Boyle DP, McArt DG, Irwin G, Wilhelm-Benartzi CS, Lioe TF, Sebastian E, et al. The prognostic significance of the aberrant extremes of p53 immunophenotypes in breast cancer. *Histopathology*. 2014;65(3):340–52.
47. Nagel JH, Peeters JK, Smid M, Sieuwerts AM, Wasieleski M, de Weerd V, et al. Gene expression profiling assigns CHEK2 1100delC breast cancers to the luminal intrinsic subtypes. *Breast Cancer Res Treat*. 2012;132(2):439–48.
48. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418–23.
49. Cheang MC, Martin M, Nielsen TO, Prat A, Voduc D, Rodriguez-Lescure A, et al. Defining breast cancer intrinsic subtypes by quantitative receptor expression. *Oncologist*. 2015;20(5):474–82.
50. Kriege M, Hollestelle A, Jager A, Huijts PE, Berns EM, Sieuwerts AM, et al. Survival and contralateral breast cancer in CHEK2 1100delC breast cancer patients: impact of adjuvant chemotherapy. *Br J Cancer*. 2014;111(5):1004–13.



51. Huzarski T, Cybulski C, Wokolorczyk D, Jakubowska A, Byrski T, Gronwald J, et al. Survival from breast cancer in patients with CHEK2 mutations. *Breast Cancer Res Treat.* 2014;144(2):397–403.
52. Cybulski C, Wokolorczyk D, Huzarski T, Byrski T, Gronwald J, Gorski B, et al. A deletion in CHEK2 of 5,395 bp predisposes to breast cancer in Poland. *Breast Cancer Res Treat.* 2007;102(1):119–22.
53. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2\*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol.* 2008;26(4):542–8.
54. Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;4(5):3.
55. Anastassiou D, Rumjantseva V, Cheng W, Huang J, Canoll PD, Yamashiro DJ, et al. Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer.* 2011;11:529.
56. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, et al. Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene.* 2006;25(15):2273–84.
57. Jechlinger M, Grunert S, Tamir IH, Janda E, Ludemann S, Waerner T, et al. Expression profiling of epithelial plasticity in tumor progression. *Oncogene.* 2003;22(46):7155–69.
58. Boquest AC, Shahdadfar A, Fronsdaal K, Sigurjonsson O, Tunheim SH, Collas P, et al. Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. *Mol Biol Cell.* 2005;16(3):1131–41.
59. Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res.* 2006;66(10):5278–86.
60. Sung SY, Hsieh CL, Law A, Zhau HE, Pathak S, Multani AS, et al. Coevolution of prostate cancer and bone stroma in three-dimensional coculture: implications for cancer growth and metastasis. *Cancer Res.* 2008;68(23):9996–10003.
61. Inoue Y, Kitagawa M, Taya Y. Phosphorylation of pRB at Ser612 by Chk1/2 leads to a complex between pRB and E2F-1 after DNA damage. *EMBO J.* 2007;26(8):2083–93.
62. Stevens C, Smith L, La Thangue NB. Chk2 activates E2F-1 in response to DNA damage. *Nat Cell Biol.* 2003;5(5):401–9.
63. Krejci P, Aklonis A, Kaucka M, Sevcikova E, Prochazkova J, Masek JK, et al. Receptor tyrosine kinases activate canonical WNT/beta-catenin signaling via MAP kinase/LRP6 pathway and direct beta-catenin phosphorylation. *PLoS One.* 2012;7(4):e35826.
64. Incassati A, Chandramouli A, Eelkema R, Cowin P. Key signaling nodes in mammary gland development and cancer:  $\beta$ -catenin. *Breast Cancer Res.* 2010;12(6):213.
65. Pond AC, Herschkowitz JI, Schwertfeger KL, Welm B, Zhang Y, York B, et al. Fibroblast growth factor receptor signaling dramatically accelerates tumorigenesis and enhances oncoprotein translation in the mouse mammary tumor virus-Wnt-1 mouse model of breast cancer. *Cancer Res.* 2010;70(12):4868–79.
66. Pond AC, Bin X, Batts T, Roarty K, Hilsenbeck S, Rosen JM. Fibroblast growth factor receptor signaling is essential for normal mammary gland development and stem cell function. *Stem Cells.* 2013;31(1):178–89.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

